

# arXiv: 25-vuotiaan vauhti kiihtyy

POSTED ON [25.10.2016](#) BY [SAMU KYTÖJOKI](#)



**VERKKARI**  
HELSINGIN YLIOPISTON KIRJASTO

[arXiv](#) on kaikkien tuntema urauurtava Open Access -arkisto ja samalla monelle suuri tuntematon. Elokuussa 25 vuotta täyttäneen julkaisuarkiston ulkokuori näyttää huomiota herättämättömän 1990-lukulaiselta. Koruttomasta ulkoasusta huolimatta sen käyttö on ollut jyrkässä kasvussa. Joka kuukausi palveluun ladataan jo noin 10 000 uutta artikkelia. arXiv ohitti viime vuonna miljoonan tallennetun artikkelin rajan, ja artikkelilatauksia on vuodesta 1994 lähtien tilastoitu 803 miljoonaa.



Palvelun perusti vuonna 1991 teoreettisen hiukkasfysiikan tutkija **Paul Ginsparg** muille saman alan tutkijoille. Ideana oli synnyttää uutuusseuranta nopeasti kehittyville fysiikan aloille ja internetin mahdollistama preprintien sähköinen jakelu. Palvelun kasvettua odotettua suuremmaksi sen majoitti Cornellin yliopiston kirjasto vuodesta 2000 eteenpäin. Ginsparg jatkoi kuitenkin palvelun pääkäyttäjänä. Tässä roolissa hän toimii edelleen.

Tutkijat näkevät arXivista sen sisällöt. Fysiikassa, tähtitieteessä, matematiikassa ja yhä kasvavassa määrin myös tietojenkäsittelytieteessä arXivin avaaminen on kuin päivän lehti: Juuri arXiviin ladattua aineistoa, josta ei ehkä ole vielä ehditty uutisoida mediassa, voi lukea tuoreeltaan. Monelle kustantajallekin arXivin tarjoama näkyvyys on plussaa. Näin ollen luonnontieteen alojen tiedekustantajia ei pääsääntöisesti haittaa, että artikkelit ovat ehtineet kerätä huomiota jo ennen ilmestymistään. Sen sijaan Naturen kaltaiset yleislehdet haluaisivat mielellään pitää uutishuomion itsellään ja näkevät arXivin kilpailijana.

Tieteenalojen lisääntymistä arXivissa rajoittavat monet käytännön syyt. Oletuksena on artikkelien laatiminen LaTeX-tekstinladontaohjelmalla, jota paljon kaavoja käyttävät tieteenalat käyttävät. Myös lingvistit laativat tekstejään samalla ohjelmalla, mutta uuden tieteenalan tuominen arXiviin vaatisi myös alaan perehtyneiden *moderaattorien* löytämistä. Artikkeleita ei voi julkaista vapaasti, vaan ne käyvät lävitse moderointiprosessin, joka on osin automaattinen ja osin vapaaehtoisverkoston pyörittämä. Kyse ei ole vertaisarvioinnista, sillä moderaattori lähinnä tarkistaa onko kyse tieteellisestä tekstistä ja onko se asiallinen.



Cornell University  
Library

We gratefully acknowledge support from  
the Simons Foundation  
and University of Helsinki

arXiv.org

Search or Article-id
 
 (Help | Advanced search)

Login

All papers

Open access to 1,196,023 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

Subject search and browse:

29 Jun 2016: [View the key findings of the arXiv user survey](#)

25 Jan 2016: [A project update, including a brief summary of activities in 2015, has been posted](#)

See cumulative "What's New" pages. Read [robots beware](#) before attempting any automated download

## Physics

- [Astrophysics \(astro-ph new, recent, find\)](#)  
includes: [Astrophysics of Galaxies](#); [Cosmology and Nongalactic Astrophysics](#); [Earth and Planetary Astrophysics](#); [High Energy Astrophysical Phenomena](#); [Instrumentation and Methods for Astrophysics](#); [Solar and Stellar Astrophysics](#)
- [Condensed Matter \(cond-mat new, recent, find\)](#)  
includes: [Disordered Systems and Neural Networks](#); [Materials Science](#); [Mesoscale and Nanoscale Physics](#); [Other Condensed Matter](#); [Quantum Gases](#); [Soft Condensed Matter](#); [Statistical Mechanics](#); [Strongly Correlated Electrons](#); [Superconductivity](#)
- [General Relativity and Quantum Cosmology \(gr-qc new, recent, find\)](#)
- [High Energy Physics - Experiment \(hep-ex new, recent, find\)](#)
- [High Energy Physics - Lattice \(hep-lat new, recent, find\)](#)
- [High Energy Physics - Phenomenology \(hep-ph new, recent, find\)](#)
- [High Energy Physics - Theory \(hep-th new, recent, find\)](#)
- [Mathematical Physics \(math-ph new, recent, find\)](#)

arXiv:n

etusivu 2016

Usein ajatellaan, että arXivissa on lähinnä fysiikan aineistoa, sillä aivan alkuvaiheessa palvelun sisältö oli lähes puhdasta hiukkasfysiikkaa. Seuraavaksi arXiviin löysivät muut fysiikan alat ja tähtitiede, sen jälkeen matematiikka. Tällä hetkellä tietojenkäsittelytiede on tieteenaloista nopeimmin kasvava. Myös tilastotiede on kasvattanut omaa osuuttaan.

Myös kirjastoammattilaisille arXivissa on oma alueensa tietojenkäsittelytieteen alla, Digital Libraries (tunnus cs.DL). [Sen uutuusvirrasta](#) löytyy muun muassa uusia bibliometriikan julkaisuja.

# Helsingin yliopiston tutkijat arXivissa

Helsingin yliopiston [TUHAT-tutkimustietokannassa](#) arXiv on monelle kirjaston julkaisutarkastajalle tuttu. Varsinkin matemaatikot ilmoittavat mielellään artikkeleittensa arXiv-versioita julkaisurekisteriin. Valtakunnallisen julkaisutiedonkeruun kriteereitä eivät arXiv-tekstit täytä – arXiv ei ole lehti, vaikka se sellaiseksi joskus paremman puutteessa kirjataan.

Minkä verran Helsingin yliopiston tutkijoiden tekstejä löytyy arXivista? Suoria lataustilastoja ei ole saatavilla, eikä arXivissa ole kirjoittajien affiliaatitietoja. Erään tarkastelun mukaan jopa puolet Helsingin yliopiston matematiikan vertaisarvioituista artikkeleista löytyy sieltä. Jos aineistoa etsii tähtitieteeseen painottuvan [NASA ADS -tietokannan](#) affiliaatio- ja kokotekstihauulla, näyttäisi siltä että vähintään neljäsosa fysiikan artikkeleistamme on arXivissa. Fysiikka kuuluu yliopiston ahkerimmin julkaiseviin aloihin, joten nämä lukumäärät ovat suuria.

## Kuinka open access arXiv on?

[arXivin oma peruslisenssi](#) on ajalta ennen Creative Commons -lisenssejä, mutta kirjoittaja voi määritellä artikkelilleen myös muun haluamansa lisenssin. Metadata on tietoisesti pidetty kevyenä ja tutkijaystävällisenä, eikä kirjoittaja aina vaivaa päätään julkaisuvaihetiedon kirjaamisella. Mikäli TUHAT-tarkastaja haluaa käyttää arXivia OA-tietojen tallennuksen lähteenä, voi olla vaikea selvittää, onko kyseessä submitted- vai accepted-vaiheen teksti. Tähtitieteessä tämä on hiukan selvempää, sillä alan lehdet suosivat embargoja.

Monet tutkijat kannattavat huokeaa ja avointa julkaisemista niin sanotuissa overlay-lehdissä. Tällöin artikkelit ovat avoimesti saatavilla arXivissa, ja lehden toimitus hyväksyy ne lehteen linkittämällä arXiv-artikkelit lehden sivuille toimitetun esittelytekstin kera. Tällaisia lehtiä on ollut matematiikassa, mutta valitettavasti ne ovat olleet toistaiseksi lyhytikäisiä. Esimerkin voi löytää TUHATista, sillä myös oman yliopistomme matematiikkaa on julkaistu Discrete Analysis -lehdessä.

On myös perinteisiä lehtikustantajia, joiden suhtautuminen arXiviin on luontevaa. IOP:n joidenkin lehtien kirjoittajaohjeissa saatetaan opastaa tutkijaa tallentamaan vertaisarvioijaa varten pelkän arXiv-tunnisteen.

arXiv on sitä käyttävien alojen tutkijoille tärkeä julkaisuväline. Uudelle käyttäjälle sen käyttöönotto saattaa olla konstikasta. arXivissa on oma viittauskäytäntönsä, jonka tulisi kuulua kirjaston viitteidenhallinnan opetukseen. Nuoret tutkijat puolestaan tarvitsevat eväitä julkaisemisprosessiin. Kuka tahansa ei voi lähettää arXiviin tekstiä, vaan nykyään uudet kirjoittajat tarvitsevat suosittelijan. Suosittelijaksi pääsee, mikäli on julkaissut arXivissa riittävästi artikkeleita tietyllä alalla. Esimerkiksi matemaattis-luonnontieteellisen tiedekunnan dekaani, matemaatikko **Jouko Väänänen** on hyväksytty suosittelijaksi matemaattisessa logiikassa.

arXivissa jokainen kirjoittaja saa oman tunnisteen, jonka voi nykyisin kytkeä [ORCID-tunnisteeseen](#). Kirjoittajien huolena on myös identifioida arXivissa omat artikkelinsa silloin, kun joku muu kirjoittajista on tallettanut ne palveluun. Kirjoittajan omalla vastuulla on tietää, voiko artikkeliversion tallettaa arXiviin vai ei, sillä mitään linkitystä Sherpa Romeon värikoodeihin tai muuta vastaataa ohjeistusta ei ole käyttäjän tukena.

## Suomeen arXiv-konsortio?

arXivin pyörittäminen maksaa vuositasolla satoja tuhansia dollareita. Open Access tavoittelee maksuttomuutta, mutta näin valtavan arkiston ylläpito on kaukana ilmaisesta. Hintalapun takaa löytyy muun muassa työvoimakustannuksia, samalla kun valtavat tietomassat vaativat palvelin- ja tiedonsiirtokapasiteettia. Cornellin yliopiston kirjastossa kukaan ei työskentele arXivin parissa kokopäiväisesti, vaan vain prosentteja kokonaistyöajasta.

Merkittävin tulolähde on vuodesta 2013 alkaen ollut arXivin käyttäjäyliopistoilta kerätyt tukimaksut. Ne on porrastettu käyttöaktiivisuuden mukaan, niin että 200 aktiivisimmalle käyttäjälle on määritelty maksuportaita. Lähes kaikki tukijoista osallistuvat kirjastojensa kautta. Suomalaisista yliopistoista Helsingin yliopisto on ollut mukana [arXivin tukiohjelmassa](#) kirjaston 2500 dollarin vuosittaisella tukimaksulla.

Helsingin yliopistossa arXivin aineistoista ladataan eniten hiukkasfysiikkaa, tähtitiedettä ja matematiikkaa. Suomessa arXivia käyttävät aktiivisesti myös Aalto-yliopisto, jossa arXivia käyttävät eniten materiaali- ja muut fyysikot sekä tietojenkäsittelytieteilijät. Jyväskylässä seurataan eniten hiukkasfysiikkaa, Oulussa tähtitiedettä ja tietojenkäsittelytiedettä. Turussa tähtitiede kiinnostaa eniten.

Aalto-yliopisto ei ole vielä mukana arXivin tukiohjelmassa vaikka se onkin 200 aktiivisimman käyttäjän joukossa. Suomalaisen melko aktiivisen käytön huomioiden voisi kysyä, kannattaisiko suomalaisten yliopistojen muodostaa aktiivisimpien arXiv-käyttäjien tukikonsortio? Näin on menetelty esim. Saksassa, Japanissa ja Iso-Britanniassa.



arXivin

Member Advisory Board koolla Ithacassa syyskuussa 2016. Cornellin kirjaston kokoushuoneesta on hulpeat näkymät Cayuga-järvelle.

arXiv tulee lähivuosina tarvitsemaan kaiken saamansa lisärahoituksen. Jo aineiston määrän kasvu vaatii panostusta, samalla kun työvoimakustannukset ovat nousussa. Palvelu tarvitsee uudistamista mm. ohjelmistopuolella, sillä nykyiset perl-skriptit eivät palvele loputtomiin. Jotta moderaattorit jaksavat käsitellä artikkeleiden suurta määrää, he tarvitsevat ajanmukaiset moderaatiovälineet joissa osa prosessista on automatisoitu.

arXiv toteutti keväällä 2016 käyttäjäkyselyn, jolla selvitettiin, mitä käyttäjät toivovat palvelun tulevaisuudelta. Vastauksia tuli hätkähdyttävät 36 000. Vastausten jakauma vastasi varsin hyvin arXivin lataustilastojen maajakautumaa .

Kyselytulosten perusteella arXivin käyttäjät ovat pääosin nuoria tutkijoita. He arvostavat kaikkein eniten palvelun ajantasaisuutta, ts. artikkelien nopeaa ilmestymistä. Peräti 95% vastaajista oli joko tyytyväisiä tai hyvin tyytyväisiä nykyiseen arXiviin. Toivotuimpiin uudistuksiin kuului mahdollisuus tallettaa artikkeliin liittyviä aineistoja, mm. dataa. Sen sijaan kommentointimahdollisuuksiin ja metriikkaan suhtauduttiin varauksellisesti.

arXivissa julkaistun aineiston saama huomio näkyy hyvin suosituimmissa altmetriikkapalveluissa, edellyttäen että artikkelin arXiv-versiota tarkastellaan erikseen. On enemmänkin sääntö kuin poikkeus, että arXiv-julkaisu saa selvästi enemmän altmetriikkahuomiota kuin lopullinen lehtiartikkeli. [Almetric.comin Top 100 artikkelista](#) 2015 peräti 5% oli ilmestynyt arXivissa. Näkyvyyspalveluita kehitettäessä olisi hyvä saada talteen myös tätä kautta kertyvä huomio.

Lue myös [arXivin laajempi esittely SlideSharessa](#).

Teksti

**Eva Isaksson**

*tietoasiantuntija*

*arXiv Member Advisory Boardin jäsen 2013-2016*